AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequences of Microviruses Associated with *Coptotermes formosanus*

Kara Schmidlin,[a,b] Simona Kraberger,[a] Rafaela S. Fontenele,[a,b] Francesca De Martini,[b] Thomas Chouvenc,[d] Gillian H. Gile,[a,b] Arvind Varsani[a,b,c,e]

[a]The Biodesign Center for Fundamental and Applied Microbiomics, Arizona State University, Tempe, Arizona, USA
[b]School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[c]Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[d]Department of Entomology and Nematology, Fort Lauderdale Research and Education Center, University of Florida, Davie, Florida, USA
[e]Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, Cape Town, South Africa

ABSTRACT Termites have a unique ability to effectively digest lignocellulose with the help of mutualistic symbionts. While gut bacteria and protozoa have been relatively well characterized in termites, the virome remains largely unexplored. Here, we report two genomes of microviruses (termite-associated microvirus-1 [TaMV-1] and termite-associated microvirus-2 [TaMV-2]) associated with the gut of *Coptotermes formosanus*.

The Formosan subterranean termite *Coptotermes formosanus* is native to China but is invasive in various subtropical areas around the world. It is an economically important species that forms large colonies and causes extensive damage to a variety of wood types (1, 2). In order to break down lignocellulose of woody plants and acquire essential nutrients, termites rely on a diverse range of hindgut symbionts, including bacteria and protozoa (3, 4). While the relationship between termites and their symbiotic gut community has been examined, the viral community remains largely unknown. Recently, 13 novel bacteriophages associated with *C. formosanus* and four novel genomoviruses with fungus-farming termites (*Odontotermes* spp.) were identified (5–7). To further characterize termite viruses, 10 *C. formosanus* gut samples were collected, pooled, and homogenized in 200 μl SM buffer (100 mM NaCl, 8 mM Mg_2SO_4, 0.01% gelatin, 50 mM Tris-HCl; Teknova, USA). The homogenate was used for viral DNA extraction, as previously described (8–10). Circular molecules were enriched by rolling circle amplification using TempliPhi 100 amplification (GE Healthcare, USA), and the resulting DNA was used to construct a 2 $\times$ 150-bp library using the Illumina TruSeq Nano DNA library prep kit and sequenced on an Illumina HiSeq 4000 platform at Macrogen, Inc. (South Korea). The raw paired-end reads (36,773,486 in total) were trimmed using Trimmomatic (11) and then *de novo* assembled using metaSPAdes 3.11.1 (12), with k-mer values of 33, 55, and 77. In the resulting 102,367 contigs (N_{50}, 1,491 nucleotides [nt]), a 4,975-nt contig (with 176$\times$ coverage) and a 4,714-nt contig (with 66$\times$ coverage) were identified as having similarities to microvirus sequences using BLASTx (13). Microviruses are prokaryote-infecting viruses with small circular single-stranded DNA genomes (14) that are packaged in icosahedral capsids (15). Within the family *Microviridae*, there are two subfamilies, *Bullavirinae,* whose members infect mainly *Enterobacteria,* and *Gokushovirinae,* whose members infect obligate intracellular parasitic bacteria (16). The genomes of termite-associated microvirus-1 (TaMV-1; GenBank accession number MH931003) and termite-associated microvirus-2 (TaMV-2; GenBank accession number MH931004) have genome organizations similar to those of other gokushoviruses (Fig. 1A and B), and phylogenetic analysis

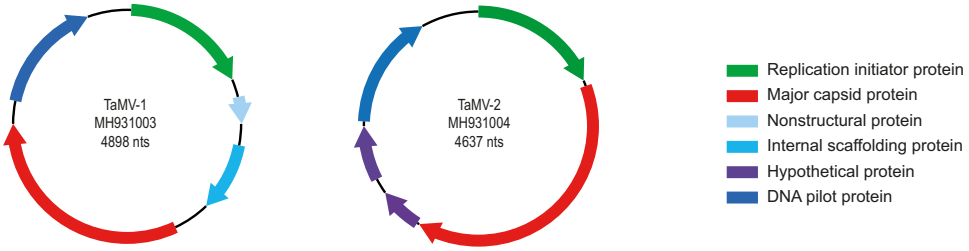
Citation Schmidlin K, Kraberger S, Fontenele RS, De Martini F, Chouvenc T, Gile GH, Varsani A. 2019. Genome sequences of microviruses associated with *Coptotermes formosanus*. Microbiol Resour Announc 8:e00185-19. https://doi.org/10.1128/MRA.00185-19.

Editor Julia A. Maresca, University of Delaware

Address correspondence to Arvind Varsani, arvind.varsani@asu.edu.

Received 20 February 2019
Accepted 25 March 2019
Published 18 April 2019

A



B

Genome Accession	Open reading frame	Blast hit	Virus name	E-value	Pairwise Identity	Coverage	Accession Number
MH931003	DNA pilot protein	DNA pilot protein	Microviridae sp. isolate ctbh687	$9x10^{-12}$	33%	83%	MH617729
	Internal scaffolding protein	Internal scaffolding protein	Microviridae sp. isolate ctdc182	$1x10^{-21}$	36%	96%	MH617685
	Major capsid protein	Major capsid protein	Eel River basin pequenovirus	0	60%	100%	KP087949
	Nonstructural protein	Nonstructural protein	Microviridae sp. isolate ctba649	$2x10^{-8}$	34%	100%	MH616777
	replication initiator protein	Replication initiator protein	Microviridae sp. isolate SD_MC_61	$3x10^{-106}$	54%	100%	MH572417
MH931004	DNA pilot protein	Putative DNA pilot protein	Microviridae sp. isolate ctcb897	$2x10^{-27}$	34%	97%	MH617732
	Hypothetical protein	Hypothetical protein	Microviridae sp. isolate ctci916	2.6	45%	27%	MH617635
	Major capsid protein	Major capsid protein	Wastewater_Microviridae_FL16	0	48%	99%	KX259470
	Hypothetical protein	Hypothetical protein	Microviridae sp. isolate SD_MF_12	0.016	36%	88%	MH572486
	Replication initiator protein	Replication initiator protein	Microviridae sp. isolate ctcd39	$3x10^{-29}$	28%	89%	MH622932

C

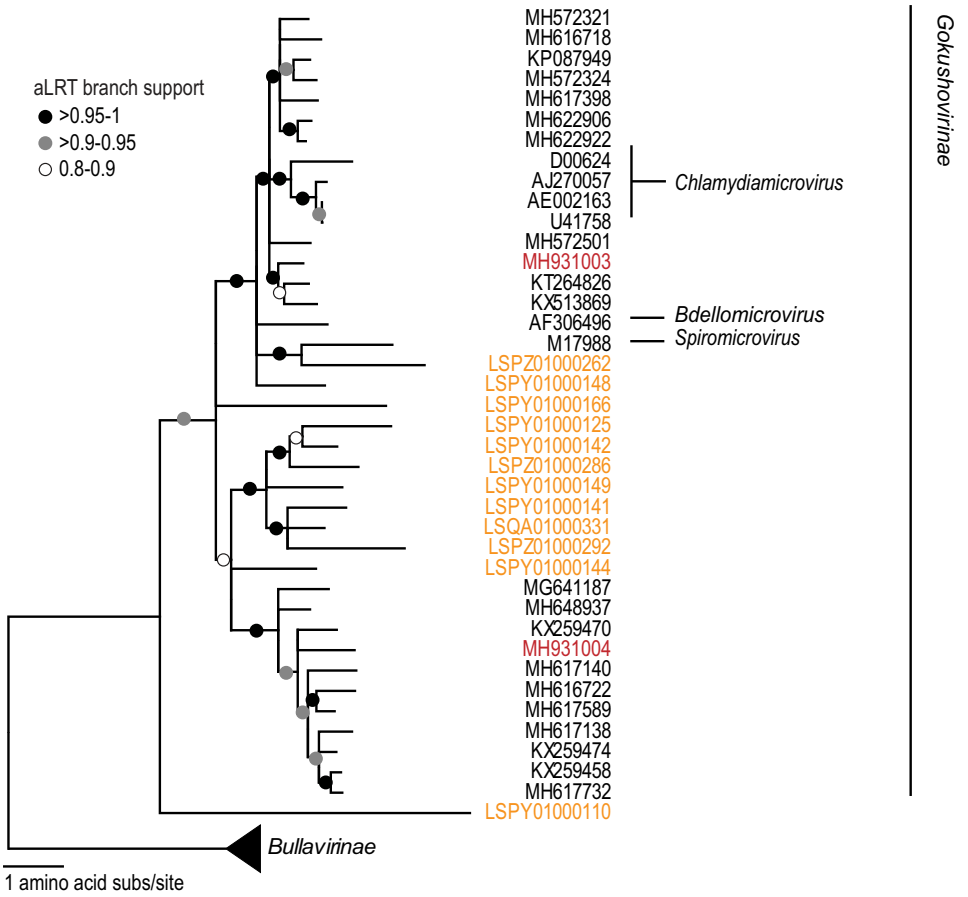


FIG 1 (A) Genome organization of termite-associated microvirus-1 (replication initiator protein, 882 nucleotides [nt]; nonstructural protein, 276 nt; internal scaffolding protein, 468 nt; major capsid protein, 1,704 nt; and DNA pilot protein, 837 nt) and termite-associated microvirus-2 (replication initiator protein, 1,017 nt; major capsid protein, 1,608 nt; hypothetical proteins, 339 and 417 nt; and DNA pilot protein, 768 nt). (B) Summary of the best BLASTp results for each ORF of TaMV-1 and TaMV-2. (C) Maximum likelihood phylogenetic tree of the MCP amino acid sequences and the pairwise identities of the MCP of most closely related *Gokushovirinae* members, those from termite reported by Tikhe and Husseneder (5), and those from this study. Numbers in red are MCP sequences from this study, and numbers in orange are MCP sequences identified in termites by Tikhe and Husseneder (5). The maximum likelihood phylogenetic trees were inferred with PhyML (17) with the RtRev+F+G substitution model and with approximate likelihood ration test (aLRT) branch support.

of the major capsid protein (MCP) confirms that both microviruses group with other members of this subfamily (Fig. 1C). TaMV-1 MCP shares ~60% amino acid identity with the MCP of the microvirus with accession number KP087949, whereas the TaMV-2 MCP shares ~48% amino acid identity with the MCP of the microvirus with accession number KX259470 (Fig. 1B). A data set of the MCPs of all published microviruses was assembled and used to query the top 10 BLASTp hits to the MCPs of TaMV-1 and TaMV-2 (Fig. 1B). These 20 MCPs, together with those from this study, those from termites reported by Tikhe and Husseneder (5), and those of classified microviruses were used to infer a maximum likelihood phylogenetic tree using PhyML (17). The MCP amino acid sequences of TaMV-1 and TaMV-2 share 36% pairwise identity with each other (Fig. 1C), with TaMV-1 clustering with MCPs of microviruses in the genus *Chlamydiamicrovirus*, whereas TaMV-2 clusters with those of unclassified microviruses. TaMV-1 and TaMV-2 are distinct from the microviruses identified by Tikhe and Husseneder (5), sharing <41% MCP amino acid identity. This highlights that there are diverse microviruses inhabiting the termite gut, and future work is needed to determine the role these viruses play in the complex host-symbiont interaction.

Data availability. The complete genome sequences of termite-associated microvirus-1 (TaMV-1) and termite-associated microvirus-2 (TaMV-2) isolates are deposited in GenBank with accession numbers MH931003 and MH931004, respectively. Raw reads have been deposited in the Sequence Read Archive (SRA) with accession number PRJNA521362.

ACKNOWLEDGMENTS

This study was supported by seed funding from the Biodesign Center of Fundamental and Applied Microbiomics, Arizona State University, USA, awarded to Arvind Varsani, and the National Science Foundation under grant number DEB-1754337, awarded to Gillian H. Gile and Arvind Varsani.

REFERENCES

1. Vargo EL, Husseneder C, Grace JK. 2003. Colony and population genetic structure of the Formosan subterranean termite, *Coptotermes formosanus*, in Japan. Mol Ecol 12:2599–2608. https://doi.org/10.1046/j.1365-294X.2003.01938.x.
2. Bourguignon T, Lo N, Šobotník J, Sillam-Dussès D, Roisin Y, Evans TA. 2016. Oceanic dispersal, vicariance and human introduction shaped the modern distribution of the termites *Reticulitermes*, *Heterotermes* and *Coptotermes*. Proc Biol Sci 283:20160179. https://doi.org/10.1098/rspb.2016.0179.
3. Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. Nat Rev Microbiol 12:168–180. https://doi.org/10.1038/nrmicro3182.
4. Evans TA, Forschler BT, Grace JK. 2013. Biology of invasive termites: a worldwide review. Annu Rev Entomol 58:455–474. https://doi.org/10.1146/annurev-ento-120811-153554.
5. Tikhe CV, Husseneder C. 2017. Metavirome sequencing of the termite gut reveals the presence of an unexplored bacteriophage community. Front Microbiol 8:2548. https://doi.org/10.3389/fmicb.2017.02548.
6. Kerr M, Rosario K, Baker CCM, Breitbart M. 2018. Discovery of four novel circular single-stranded DNA viruses in fungus-farming termites. Genome Announc 6:e00318-18. https://doi.org/10.1128/genomeA.00318-18.
7. Tikhe CV, Martin TM, Gissendanner CR, Husseneder C. 2015. Complete genome sequence of *Citrobacter* phage CVT22 isolated from the gut of the formosan subterranean termite, *Coptotermes formosanus* Shiraki. Genome Announc 3:e00408-15. https://doi.org/10.1128/genomeA.00408-15.
8. Kraberger S, Polston JE, Capobianco HM, Alcalá-Briseño RI, Fontenele RS, Varsani A. 2017. Genomovirus genomes recovered from *Echinothrips americanus* sampled in Florida, USA. Genome Announc 5:e00445-17. https://doi.org/10.1128/genomeA.00445-17.
9. Waits K, Edwards MJ, Cobb IN, Fontenele RS, Varsani A. 2018. Identification of an anellovirus and genomoviruses in ixodid ticks. Virus Genes 54:155–159. https://doi.org/10.1007/s11262-017-1520-5.
10. Kamali M, Heydarnejad J, Pouramini N, Masumi H, Farkas K, Kraberger S, Varsani A. 2017. Genome sequences of *Beet curly top Iran virus*, *Oat dwarf virus*, *Turnip curly top virus*, and *Wheat dwarf virus* identified in leafhoppers. Genome Announc 5:e01674-16. https://doi.org/10.1128/genomeA.01674-16.
11. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
14. Doore SM, Fane BA. 2016. The microviridae: diversity, assembly, and experimental evolution. Virology 491:45–55. https://doi.org/10.1016/j.virol.2016.01.020.
15. Chipman PR, Agbandje-McKenna M, Renaudin J, Baker TS, McKenna R. 1998. Structural analysis of the *Spiroplasma virus*, SpV4: implications for evolutionary variation to obtain host diversity among the *Microviridae*. Structure 6:135–145. https://doi.org/10.1016/S0969-2126(98)00016-1.
16. Roux S, Krupovic M, Poulet A, Debroas D, Enault F. 2012. Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One 7:e40418. https://doi.org/10.1371/journal.pone.0040418.
17. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.